Structural Insights into Substrate Specificity in Variants of N-Acetylneuraminic Acid Lyase Produced by Directed Evolution

Ivan Campeotto¹, Amanda H. Bolt¹,², Thomas A. Harman¹,², Caitriona Dennis¹, Chi H. Trinh¹, Simon E. V. Phillips¹, Adam Nelson¹,², Arwen R. Pearson¹ and Alan Berry¹∗

¹Astbury Center for Structural Molecular Biology, Garstang Building, University of Leeds, Leeds LS2 9JT, UK
²School of Chemistry, University of Leeds, Leeds LS2 9JT, UK

The substrate specificity of Escherichia coli N-acetylneuraminic acid lyase was previously switched from the natural condensation of pyruvate with N-acetylmannosamine, yielding N-acetylneuraminic acid, to the aldol condensation generating N-alkylcarboxamide analogues of N-acetylneuraminic acid. This was achieved by a single mutation of Glu192 to Asn. In order to analyze the structural changes involved and to more fully understand the basis of this switch in specificity, we have isolated all 20 variants of the enzyme at position 192 and determined the activities with a range of substrates. We have also determined five high-resolution crystal structures: the structures of wild-type E. coli N-acetylneuraminic acid lyase in the presence and in the absence of pyruvate, the structures of the E192N variant in the presence and in the absence of pyruvate, and the structure of the E192N variant in the presence of pyruvate and a competitive inhibitor (2R,3R)-2,3,4-trihydroxy-N,N-dipropylbutanamide. All structures were solved in space group P2₁ at resolutions ranging from 1.65 Å to 2.2 Å. A comparison of these structures, in combination with the specificity profiles of the variants, reveals subtle differences that explain the details of the specificity changes. This work demonstrates the subtleties of enzyme–substrate interactions and the importance of determining the structures of enzymes produced by directed evolution, where the specificity determinants may change from one substrate to another.

© 2010 Elsevier Ltd. Open access under CC BY license.
Introduction

N-Acetylenuraminic acid lyase (NAL; EC 4.1.3.3, otherwise known as N-acetylenuraminic acid aldolase or sialic acid aldolase) catalyzes the reversible aldol condensation of N-acetylmannosamine (ManNAc) with pyruvate to yield the most abundant sialic acid N-acetylenuraminic acid (Neu5Ac) (Fig. 1). Sialic acids are essential components of complex carbohydrates that play pivotal roles as recognition signals in a variety of biological processes. Such molecular recognition events are well exemplified in host–pathogen and host–parasite interactions where the oligosaccharide is often required for the invasion, infectivity, and survival of the invading organism in the host (for a review of the roles of sialic acids, see Rosenberg). Sialic acid mimetics can, therefore, be important chemotherapeutic agents (e.g., the anti-influenza drugs Relenza and Tamiflu). NAL has been used in a number of syntheses of such analogues; however, its synthetic utility has been limited by its substrate specificity (e.g., two-carbon and three-carbon aldehyde acceptors are very poor substrates) and by the poor facial stereoselectivity of wild-type NAL and other members of the NAL superfamily during carbon–carbon bond formation. To overcome these problems, we have used structure-guided protein engineering and directed evolution of the Escherichia coli enzyme to create a pair of complementary biocatalysts to synthesize either possible diastereoisomeric aldol product or to tailor the enzyme for the parallel synthesis of sialic acid mimetics. We have reported that the engineered NAL variant E192N catalyzes the reversible cleavage of N-dialkylcarboxamide analogues of Neu5Ac—such as (5R,6R)-7-(dipropylamino)-4,5,6-trihydroxy-2,7-dioxoheptanoic acid (DPAH) into pyruvate and (2R,3S)-2,3-dihydroxy-4-oxo-N,N-dipropylbutanamide (DHOB) (Fig. 1)—five times more effectively that the wild-type enzyme catalyzes the cleavage of Neu5Ac. The residue responsible for this alteration of specificity at position 192 lies at one end of the active-site cleft opposite the active-site catalytic residue Lys165, which forms a Schiff base with the substrate as part of the catalytic cycle (Fig. 2a). In order to probe the mechanism by which alteration of the residue at position 192 affects substrate specificity, we have undertaken an analysis of the activity of all 20 variants of NAL at position 192 along with an X-ray structural analysis of the E192N enzyme in the absence and in the presence of a competitive inhibitor, (2R,3R)-2,3,4-trihydroxy-N,N-dipropylbutanamide (THB) (Fig. 1). These studies have provided new insights into the binding of substrates to the wild-type and variant enzymes and suggest the structural basis for the discrimination of substrates by the E192X variants.

Results and Discussion

The importance of various interactions between wild-type NAL and different substrates or substrate analogues has been revealed by an analysis of the substrate specificity of wild-type NAL and by X-ray crystal structures. In particular, the structures of E. coli NAL in the absence of any substrate, the hydroxypyruvate or sodium-borohydride-reduced pyruvate complexes, and Haemophilus influenzae NAL in the absence of substrate or in the presence of sialic acid alditol (Neu5Ac2ol), 4-deoxysialic acid (4d-Neu5Ac), or 4-oxo-sialic acid (4oxo-Neu5Ac) have revealed that the residues at positions 191, 192, and 208 (E. coli numbering is used throughout this article) interact strongly with the hydroxyl groups at C7, C8, and C9 of Neu5Ac (Fig. 2). Saturation mutagenesis of these residues in the E. coli enzyme allowed us to identify residue 192 as crucial in switching the substrate specificity from Neu5Ac to DPAH.

Activity profile of NAL variants at position 192

In order to determine the relationship between the amino acid at position 192 and the activity with various compounds, we sequenced 96 clones from an E192X saturation mutagenesis library. This allowed us to identify 15 of the 20 possible variants of NAL at position 192. The remaining variants (E192A, E192F, E192H, E192Q, and E192Y) were then constructed by site-directed mutagenesis. All variants of the E192X library were expressed in E. coli as hexa-His-tagged proteins and purified to homogeneity on Ni-NTA resin. The activity of each enzyme for the cleavage of either Neu5Ac or DPAH was determined at a single low concentration of substrate (0.5 mM) so that the enzymes were unlikely to be saturated (the Km of the wild-type enzyme for Neu5Ac is ~4 mM), and the observed rates may be an approximation of kcat/Km (Fig. 3). These results show that the activity towards Neu5Ac is highest for the wild-type enzyme (E192), and only the E192Q variant shows significant activity towards this substrate. The steady-state kinetic parameters kcat and Km were then determined for the cleavage of Neu5Ac with several NAL variants, as shown in Table 1. These results show that correct interaction between Neu5Ac and residue 192 is extremely important for binding and catalysis. In the X-ray crystal structure of H. influenzae NAL in complex with 4-oxo-Neu5Ac, residue Glu192 hydrogen bonds with the diol at C8 and C9 of the sialic acid analogue (Fig. 2), and it is likely that the glutamine introduced in E192Q mimics this interaction so well that the kinetic parameters of the E192Q enzyme with Neu5Ac (Table 1) are not significantly altered with respect to wild type. This profile of activity, with only E192...
and E192Q producing significant activity, explains the distribution of activities towards Neu5Ac found during the initial isolation of a NAL variant capable of turning over the dipropylamide DPAH. Interestingly, maintaining the hydrogen-bonding capacity at position 192 while reducing the length of the side chain (e.g., in the E192D or E192N variants) severely reduced the activity ($k_{\text{cat}}/K_m$) to less than 15% of the wild-type value with Neu5Ac (Table 1). This is presumably because the substrate, in its extended conformation, cannot span from the introduced aspartate or asparagine to lysine 165, where it has to form a Schiff base. This finding is also in line with previous studies that have shown that short carbon chain analogues of Neu5Ac are not substrates for the wild-type enzyme.

The activity profile of the E192X library with DPAH differs dramatically from that with Neu5Ac, as many more of the variants (e.g., E192F, E192H, E192M, E192N, E192Q, E192P, and E192V) show significant activity with the new substrate (Fig. 3). The steady-state kinetic parameters of these enzymes (Table 1) confirmed that the E192N enzyme had the highest activity for DPAH. It is interesting to note that E192Q shows a high level of activity with DPAH, making this enzyme the most promiscuous in terms of substrate specificity. All of the active variants showed $k_{\text{cat}}$ values generally similar to those of the wild-type enzyme, and changes in the specificity constant $k_{\text{cat}}/K_m$ were all brought about by changes in $K_m$ for the substrate (Table 1), implying alterations in the ability of the enzyme to recognize and bind the substrate depending on the diversity of functionality at position 192. The range of active functionalities at position 192 suggests that the interaction with the substrate in these enzymes is less specific than the interaction between E192 and Neu5Ac, and that a number of residues are able to make interactions with DPAH to promote catalysis. Interestingly, some of the active variants have large hydrophobic side chains at position 192 (E192F, E192Y, and E192M), and we reasoned that these side chains may provide a hydrophobic environment with which the propyl groups of DPAH are able to interact in a favorable manner. In order to gain further insights into the mechanism of substrate specificity in the E192 variants, we embarked on an X-ray crystallographic analysis of the ‘best’ enzyme for dipropylamide cleavage (namely E192N) and on a structural modeling of the possible binding modes of interaction of DPAH with other E192 variants.

**Crystallographic results**

NAL has a ($\beta/\alpha$)$_8$ barrel structure of the type first found in triose phosphate isomerase, and crystal structures of *E. coli* NAL in the absence of ligands [Protein Data Bank (PDB) code 1NAL], in complex with hydroxypropionate (PDB code 1FDY), or trapped by sodium borohydride in the presence of pyruvate (PDB code 1FDZ) were already available in the PDB database. In addition, the structure of an L142R mutant of *E. coli* NAL, which was evolved in vitro to have increased dihydrolipicolinate synthase activity, was also available in complex with hydroxypropionate (PDB code 1HL2). We reasoned that a structural understanding of the features that caused the change in the specificity of the E192N variant away from the reaction of pyruvate with ManNAc towards that with DHOB would only be forthcoming if we could solve the structure of NAL in the presence of substrates or substrate analogues that mimicked the full-length substrate. However, with the exception of the L142R mutant NAL structure, *E. coli* NAL crystals had been obtained after crystal growth in high-salt concentrations, resulting in the presence of a sulfate ion in the catalytic pocket that...
precluded co-crystallization with ligands larger than pyruvate.\textsuperscript{14,16} We therefore identified new conditions for the crystallization of \textit{E. coli} NAL (in space group $P_{21}$) that avoided the use of sulfate ions and allowed us to solve the structures of both the wild type and the E192N variant in the presence and in the absence of a number of ligands.

The crystal structures of wild-type \textit{E. coli} NAL and the E192N variant were solved at 2.2 Å and 1.8 Å, respectively (Table 2), and show a tertiary structure identical to that previously observed.\textsuperscript{14} The overall RMSD between the $C\alpha$ atoms in the wild-type structure and the $C\alpha$ atoms in the E192N structure is 0.3 Å. The only significant differences are the orientation of Asn192 compared to that of the wild-type Glu192 and the resulting changes in surface potential at the entrance of the catalytic pocket when the negatively charged glutamate is replaced (Fig. 4).

The structures of complexes with pyruvate were also obtained (to 1.65 Å resolution for the wild-type enzyme and to 1.8 Å resolution for E192N; Table 2). As expected, pyruvate was covalently linked to Lys165, and the planarity of electron density is consistent with the formation of the natural pyruvate enamine intermediate (Figs. 2a and 5). We have previously reported the structure of the E192N variant, in space group $P_{212121}$, in complex with pyruvate.\textsuperscript{19} The interactions made by pyruvate in that structure and in the $P_{21}$ structures of the wild-type–pyruvate and E192N–pyruvate complexes reported here are identical. The carboxylate group of pyruvate forms hydrogen bonds with the backbone amides of Ser47 and Thr48 and also with the side-chain OH of Thr48 (Fig. 5), as previously found for pyruvate and hydroxypyruvate\textsuperscript{15} and as is typical of a number of related pyruvate-dependent aldolases.\textsuperscript{16} Comparison with the apo-NAL structures indicates that there are no significant changes in the active site upon pyruvate binding; the only difference in the two structures appears to be a further disordering of the C-terminal residues in a remote location from the catalytic pocket.

Soaking the E192N variant with the substrate DPAH in order to trap a DPAH–NAL complex was unsuccessful due to the rapid rate of the cleavage.
reaction. As a result of this, only density for the cleavage product (pyruvate covalently linked to Lys165) was observed, and this structure was identical with the E192N–pyruvate complex described above (data not shown). To overcome this problem, we designed an analogue of the acceptor DHOB (Fig. 1) in the E192N-catalyzed reaction. This compound, THB (Fig. 1), mimics the substrate DHOB, but the reactive aldehyde has been reduced to an alcohol, preventing nucleophilic attack of pyruvate enamine. THB was synthesized, and steady-state enzyme kinetics showed that it is a competitive inhibitor of the E192N NAL reaction (Fig. 6). The $K_i$ of THB for the reaction was measured as 1.3 mM.

The structure of the E192N complex with pyruvate and THB was solved at 2.05 Å resolution and, overall, appeared identical with the structure of E192N in complex with pyruvate, with additional electron density only observed in the active site. This strong electron density corresponds to the pyruvate enamine with residue Lys165. The planarity of this electron density confirms the expected sp$^2$ hybridization for the carbon–carbon double bond of the enamine intermediate. The noncovalent interactions of pyruvate with the protein are identical with those observed for the wild-type–pyruvate and E192N–pyruvate complexes. The $F_o-F_c$ electron density map also revealed a further strong positive electron density in the enzyme active site. The inhibitor THB fits well into this density, with the exception of the terminal ends of the two propyl groups, likely due to the freedom of rotation of these groups (Fig. 7). There

Fig. 3. Activity of NAL variants isolated and purified from the E192X saturation mutagenesis library with (a) Neu5Ac (0.5 mM) and (a) DPAH (0.5 mM). Each assay was carried out as described in Materials and Methods and contained ~0.03 mg of purified enzyme.
Table 1. Steady-state kinetic parameters of wild-type and E192X variants with Neu5Ac and DPAH

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Neu5Ac</th>
<th>DPAH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_{cat}$ (min$^{-1}$)</td>
<td>$K_m$ (mM)</td>
</tr>
<tr>
<td>Wild type</td>
<td>260±6$^a$</td>
<td>4.4±0.3$^a$</td>
</tr>
<tr>
<td>E192N</td>
<td>170±10$^a$</td>
<td>38±5$^a$</td>
</tr>
<tr>
<td>E192Q</td>
<td>300±20</td>
<td>3.5±0.5</td>
</tr>
<tr>
<td>E192V</td>
<td>120±8$^a$</td>
<td>34±5$^a$</td>
</tr>
<tr>
<td>E192F</td>
<td>160±5</td>
<td>5.9±0.4</td>
</tr>
<tr>
<td>E192H</td>
<td>120±5</td>
<td>6.2±0.7</td>
</tr>
<tr>
<td>E192M</td>
<td>70±4</td>
<td>5±0.8</td>
</tr>
<tr>
<td>E192P</td>
<td>25±1</td>
<td>6±0.9</td>
</tr>
<tr>
<td>E192D</td>
<td>120±5</td>
<td>4.6±0.5</td>
</tr>
<tr>
<td>E192S</td>
<td>50±3</td>
<td>6±0.8</td>
</tr>
</tbody>
</table>

The steady-state kinetic parameters of the wild-type and variant enzymes for Neu5Ac and DPAH cleavage were measured with a standard coupled enzyme assay. Kinetic parameters ($±$ standard error of the fit) were determined by fitting the data to the Michaelis-Menten equation.

$^a$ Data taken from Williams et al.$^1$

---

Table 2. Crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>Wild type</th>
<th>Wild type+ pyruvate</th>
<th>E192N</th>
<th>E192N+ pyruvate</th>
<th>E192N+ pyruvate+ THB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diamond beamline</td>
<td>102</td>
<td>103</td>
<td>104</td>
<td>103</td>
<td>103</td>
</tr>
<tr>
<td>Space group</td>
<td>P21</td>
<td>P21</td>
<td>P21</td>
<td>P21</td>
<td>P21</td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>54.8, 142.2, 84.2</td>
<td>54.7, 142.5, 83.6</td>
<td>54.6, 142.8, 84.5</td>
<td>56.9, 143.0, 83.9</td>
<td>57.0, 143.7, 83.3</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>79.56–2.20</td>
<td>79.1–1.65</td>
<td>47.60–1.80</td>
<td>79.06–1.80</td>
<td>79.31–2.05</td>
</tr>
<tr>
<td>$R_{merge}$ (all I$^+$ and I$^-$)</td>
<td>(2.26–2.20)</td>
<td>(1.74–1.65)</td>
<td>(1.90–1.80)</td>
<td>(1.95–1.80)</td>
<td>(2.16–2.05)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100)</td>
<td>91.2 (60.8)</td>
<td>98.4 (97.2)</td>
<td>100 (99.9)</td>
<td>98.6 (96.3)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.7 (3.7)</td>
<td>3.4 (2.9)</td>
<td>3.6 (3.5)</td>
<td>3.6 (3.5)</td>
<td>3.5 (3.4)</td>
</tr>
<tr>
<td>Wilson B-factor (Å$^2$)</td>
<td>34.8</td>
<td>31.7</td>
<td>26.8</td>
<td>21.8</td>
<td>24.5</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>197,621 (29,177)</td>
<td>449,770 (37,158)</td>
<td>398,910 (55,837)</td>
<td>391,185 (54,278)</td>
<td>277,866 (37,608)</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>54,006 (7847)</td>
<td>131,694 (12,746)</td>
<td>110,985 (15,919)</td>
<td>107,356 (15,671)</td>
<td>78,597 (11,141)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>79.56–2.20</td>
<td>79.1–1.65</td>
<td>47.60–1.80</td>
<td>79.06–1.80</td>
<td>79.31–2.05</td>
</tr>
<tr>
<td>$R$-factor</td>
<td>0.211 (0.259)</td>
<td>0.209 (0.261)</td>
<td>0.197 (0.261)</td>
<td>0.188 (0.301)</td>
<td>0.190 (0.281)</td>
</tr>
<tr>
<td>$R_{free}$</td>
<td>0.271 (0.356)</td>
<td>0.247 (0.344)</td>
<td>0.253 (0.391)</td>
<td>0.225 (0.344)</td>
<td>0.234 (0.337)</td>
</tr>
<tr>
<td>Twinning fraction</td>
<td>0.338</td>
<td>0.334</td>
<td>0.463</td>
<td>None</td>
<td>0.06</td>
</tr>
<tr>
<td>Number of atoms</td>
<td>Protein</td>
<td>9166</td>
<td>9166</td>
<td>9178</td>
<td>9292</td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>169</td>
<td>758</td>
<td>316</td>
<td>654</td>
</tr>
<tr>
<td></td>
<td>Ligands</td>
<td>—</td>
<td>31 (PEG 400, Na$^+$)</td>
<td>4 (Cl)</td>
<td>24 (PEG 400, pyruvate$^b$)</td>
</tr>
<tr>
<td>Average B-factors (Å$^2$)</td>
<td>32.9</td>
<td>22.8</td>
<td>28.0</td>
<td>22.8</td>
<td>28.4</td>
</tr>
<tr>
<td></td>
<td>Waters</td>
<td>27.9</td>
<td>26.7</td>
<td>26.7</td>
<td>27.2</td>
</tr>
<tr>
<td></td>
<td>Ligands</td>
<td>—</td>
<td>32.4</td>
<td>26.3</td>
<td>45.8</td>
</tr>
<tr>
<td>RMSDs</td>
<td>Bond lengths (Å)</td>
<td>0.013</td>
<td>0.012</td>
<td>0.011</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>Bond angles (°)</td>
<td>1.46</td>
<td>1.80</td>
<td>1.33</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>Ramachandran plot, most favored region (%)</td>
<td>94.2$^b$</td>
<td>98.3$^b$</td>
<td>97.6$^a$</td>
<td>98.0$^b$</td>
</tr>
<tr>
<td>PDB code</td>
<td>2WO5</td>
<td>2WNN</td>
<td>2WNQ</td>
<td>2WNZ</td>
<td>2WPB</td>
</tr>
</tbody>
</table>

$^a$ Two additional molecules of noncovalently bound pyruvate are observed in subunits C and D.

$^b$ In all structures, 100% of residues lie within favored regions of the Ramachandran plot.
was some variation in the quality of the electron density and in the final modeled orientation for THB in the four crystallographically independent subunits of the tetramer, with subunit D showing the most well-defined THB density. The electron density in subunit B was more consistent with a slightly different binding mode for THB in comparison to subunits A, C, and D (Fig. 8) (see the text below and Table S1).

In all four subunits, H-bonds are formed between the C4 hydroxyl group of THB and the backbone amide of Ser208 and the side chain of Asp191 (Fig. 8). In addition, the backbone carbonyl group of Gly189 makes H-bonds with the THB (with the C3 hydroxyl group in subunits A, C, and D, and with the C2 hydroxyl in subunit B (equivalent to C5 and C6 of DPAH)). These interactions result in the more polar portion of the inhibitor (containing the three hydroxyl groups) being bound at the base of the catalytic pocket next to the pyruvate (Figs. 7 and 8). During enzyme catalysis, a new carbon–carbon bond is formed between C3 from the pyruvate enamine and the C4 from the aldehyde acceptor. In the enzyme–pyruvate–THB complex structure, this bond cannot be formed, as THB is unable to undergo nucleophilic attack by the enamine C3. Nonetheless, the C3 of the pyruvate enamine lies only ~3.5 Å from the C4 of THB, compatible with these two atoms being in van der Waals contact. This distance compares well with the separation of the substrate aldol donor and acceptor molecules (3.3–4.4 Å) found in the X-ray structure of a different aldolase, the *Mycobacterium*
*tuberculosis* fructose bisphosphate aldolase,\(^2\) and therefore represents the Michaelis entry (or exit) complex in which the two substrates are ideally placed, ready for the nucleophilic attack of the enamine on the aldehyde of the acceptor substrate.

The structure of E192N, in complex with pyruvate and THB, reveals the structural basis of the discrimination between substrates manifested by wild-type and E192N NAL. In the wild-type enzyme, residue Glu192 is sandwiched between Tyr190 of the same

---

**Fig. 5.** Active site of the wild-type NAL–pyruvate complex. Two views of the active site of the wild-type NAL–pyruvate complex showing the interactions of the pyruvate enamine with the enzyme. Pyruvate is covalently linked to Lys165 and makes hydrogen bonds with Ser47 and Thr48. The hydrogen-bonding network to Tyr137 and a conserved water molecule are also shown. The lower panel shows a rotation of the molecule to display the planar nature of the \(2F_o-F_c\) electron density (contoured at 1 RMSD) of pyruvate–enzyme bonding confirming the enamine structure for this intermediate.
subunit and Tyr172 of a neighboring subunit. In this position, its carboxylic acid group points towards the active-site cleft, perfectly positioned to form strong hydrogen bonds with the O8 and O9 of the glycerol moiety of the incoming substrate Neu5Ac (or the equivalent O5 and O6 of ManNAc). In contrast, in the E192N variant, the amide of Asn192 points out of the active-site cleft (a position disfavored for glutamate due to unfavorable contacts with the neighboring hydrophobic residues Tyr172 and Ile243), and this conformation produces a larger hydrophobic binding surface that is ideal for interaction with THB. The two conformations of THB (found in subunits A, C, and D compared with subunit B; Fig. 8) occur because of a rotation of the main-carbon chain of THB around the C4–C3 bond. This places the two propyl arms in different positions in the two conformations. In subunits A, C, and D, the syn arm of the dipropylamide lies against the hydrophobic surface of Tyr190, Ile247, and Tyr172 (this latter residue is from a neighboring subunit), while the anti arm lies in a deep hydrophobic pocket composed of Ile243, Thr209, Val251, and Leu246. In this binding conformation, the mutated residue at position 192 lies at the bifurcation of the two propyl arms of THB. Modeling of a glutamate residue at position 192 shows that the surface of the wild-type enzyme would be more polar in this important binding region and that there would also be steric clashes between the enzyme carboxylate and the C1 and C2 atoms of the dipropylamide arms. In subunit B, the other conformation of bound THB places the syn arm of the inhibitor in the same region as in subunits A, C, and D, on the surface of Tyr190 and Ile247; however, in this case, the hydrophobic pocket is empty, with the anti arm instead lying against another hydrophobic part of the surface formed from residues Val251 and Phe252. In this case, the introduced Asn192 lies between the syn-propyl arm and the amide carbonyl group. The modeled position of the wild-type glutamate residue again shows steric clashes of the Glu192 carboxylate, this time with both the syn-propyl and the amide O atom. With the assumption that the position and orientations of the bound competitive inhibitor THB accurately reflect the binding of the substrate DPAH, the structural basis of discrimination against DPAH binding to the wild-type enzyme and in favor of binding to the E192N variant becomes clear. The larger glutamate at position 192 in the wild-type enzyme prevents DPAH binding not only through steric clashes with the substrate, which cannot be relieved because of the planarity of the substrate, but also through the negative charge of Glu192 making unfavorable interactions with the hydrophobic parts of DPAH. An asparagine at position 192, on the other hand, provides an uncharged surface with no steric interference with DPAH binding. These results also explain the substrate activity profile found previously for the E192N variant. Dialkyl substituents, including dibutylamides, are well accepted by the E192N enzyme, and we would expect that the syn arm of each amide would bind similarly to that of THB, while the hydrophobic pocket formed from Ile243, Thr209, Val251, and Leu246 described above is deep enough to accommodate the longer butyl group. We suggest that other active substrates, including pyrrolidinyl amide, morpholinyl amide, and piperidinyl amide, interact with the E192N variant through the extensive hydrophobic surface generated by removing the charge from position 192, and this explains why the more hydrophobic cyclic substrates are preferred.

**Understanding the role of residue 192 in substrate selectivity**

Catalysis of the aldol condensation/cleavage of the natural substrates of NAL is supported by a...
narrow range of amino acids at position 192 (E and Q; Fig. 3a) characterized by their polarity and size, as described above. In contrast, the reaction with DPAH and other dialkyl derivatives tolerates a much wider range of side-chain size and polarity at this position (Fig. 3b), with E192Q showing high levels of activity with both natural and synthetic substrate classes. Our structural investigations of the E192N enzyme and, in particular, its complex with the inhibitor THB have enabled us to propose a general structural basis for these findings. The successful binding and catalysis of the reaction with dialkylamides clearly require a suitably sized hydrophobic binding site. Very large residues at position 192 (e.g., E192W) or charged residues (E192R, E192K, E192D, and E192E) would interfere with substrate binding by breaking up the hydrophobic binding surface required, and this is reflected in their low activity (Fig. 3b). On the other hand, we suggest that small residues at position 192 (E192G, E192A, and E192S) do not provide enough surface for interaction with the substrate, explaining their lack of activity. Of more interest is how larger residues at position 192 (e.g., E192F, E192Y, E192Q, E192M, E192V, and E192P) are still able to promote reasonable levels of activity when we might have expected these large residues to clash with the alkyl arms of the substrate. However, in silico mutagenesis of residue 192 revealed a number of side-chain conformations that allow packing of the mutated

**Fig. 7.** X-ray crystal structure of the E192N NAL–pyruvate–THB complex. (a) The active site of the E192N NAL enzyme in the presence of pyruvate and the competitive inhibitor THB. The green net shows a THB electron density omit map contoured at an RMSD of 1.5. Interactions of pyruvate enamine with Ser47 and Thr48, as well as the interactions of THB with Ser208 and Asp191, are also shown. (b) Details of Lys165–pyruvate enamine and THB. The green net shows the THB omit map as described above. $2F_o - F_c$ electron density contoured at an RMSD of 1.0 is shown as a blue net for the Lys165–pyruvate Schiff base. The distance (3.6 Å) between the pyruvate methyl group and the C4 of THB (equivalent to the new carbon–carbon bond formed between the substrate acceptor and the donor) is shown as a red dotted line.
side chain into the overall protein structure without significant clashes with the position of the THB bound in our crystal structure. For example, both Tyr192 and Phe192 can pack in a hydrophobic site between residues Tyr172 and Tyr190, providing a suitable hydrophobic surface for THB binding without steric clashes. Similarly, Met192 can lie in this region with the terminal methyl group rotated away from the substrate to provide room for THB binding, while smaller hydrophobic residues such as valine and proline can also fit into the folded chain at position 192 without serious distortions to the backbone of the protein and without interfering with the binding of THB and hence of the dialkylamide substrates. A similar analysis reveals the basis of the promiscuity of the E192Q enzyme (Fig. 3a and b). If modeled in the same conformation as the native glutamate, the introduced glutamine is able to form H-bonds with the O8 and O9 hydroxyl groups of the natural substrate Neu5Ac. However, this conformation of the glutamine would cause rejection of DPAH as a substrate because of steric interference. On the other hand, an alternative conformation of the neutral glutamine side chain, in which it can pack between the two active-site tyrosine residues in a conformation disfavored for the negatively charged glutamate residue, is possible. In this conformation, it provides a hydrophobic surface with no steric problems for DPAH binding. In this way, the E192Q enzyme is able to provide alternative modes of substrate binding that account for its promiscuity.

**Conclusions**

The directed evolution approach for the construction of new substrate specificities does not require any previous structural or mechanistic knowledge of catalysis in the target enzyme, and sequence analysis of the resultant successful variants selected often contains unexpected residues at unexpected positions to produce the desired activity. However, a full understanding of the mechanisms underlying the changes in specificity is required if we are to learn how to rationally produce new activities, and reports of the structures of enzymes with altered substrate
specificity created by directed evolution are appearing. These have revealed that altered substrate specificity can be brought about in a number of ways. For example, in aspartate aminotransferase, changes in residues involved in direct binding to the substrate or conformational changes in the enzyme brought about by mutations often at a distance of >10 Å from the active site have been demonstrated to affect specificity. Similarly, the promotion of a more compact conformation of the acyltransferase LovD brought about by mutations was found to be important in creating a simvastatin synthase from this template. Finally, significant changes in the size and conformation of active-site residues resulting in the reshaping of the active site were considered a major determinant of the specificity switches found in evolved enantioselective epoxide hydrolase. Here, by combining substrate specificity profiles with X-ray crystal structure determination, we have been able to dissect the way that the residue present at position 192 of NAL can affect substrate specificity. The change from Glu192 to Asn192 to switch the specificity from Neu5Ac to DPAH was not a change we would have rationally chosen to move from a triol-containing substrate to the more hydrophobic and bulkier dipropylamide. Our results have shown that when the mutations at position 192 provide an extensive hydrophobic surface capable of binding dipropyl-containing substrates, there are a number of conformations of the substrate that result in activity. However, if there is too much space between the enzyme and the substrate (e.g., with small or hydrophilic residues at position 192), then the H-bonding at the other end of the substrate is not sufficient to provide enough binding energy to hold the substrate for catalysis to occur. This accords with the lack of enzyme activity for ‘short’ two-carbon or three-carbon aldehyde substrates. These results point the way towards engineering NAL variants with activity for such substrates, perhaps by reducing the size of the active site around the aldehyde substrate binding pocket, and highlight the need for future structural studies of evolved enzymes.

Materials and Methods

Expression and purification of NAL

E. coli BL21-competent cells were transformed with the plasmid pKanA harboring the wild-type NAL gene or the E192X NAL library members, as previously described. Cells were grown in 5 mL of 2YT medium containing 50 μg mL⁻¹ ampicillin and inoculated in 2 L of 2YT medium containing 50 μg mL⁻¹ ampicillin and 0.4 mM IPTG for 16 h at 30 °C. Cells were lysed in a Cell Disruption Systems apparatus at 19,000 psi in 50 mM potassium phosphate buffer (pH 7.5) containing 300 mM NaCl supplemented with protease cocktail inhibitor (Roche). NAL was bound to Qiagen Ni-NTA resin equilibrated in the same buffer containing 30 mM imidazole and was eluted at 500 mM imidazole. Purified NAL was dialyzed overnight against 50 mM Tris·HCl and 150 mM NaCl (pH 7.5). Further purification was achieved by size-exclusion chromatography using an S-200 gel-filtration column (Amersham) in the same buffer before storage and concentration. Protein purity was assessed by SDS-PAGE, and molecular weight was confirmed with electrospray ionization mass spectrometry.

Enzyme assay

Aldolase activity with Neu5Ac and analogues was measured using a coupled enzyme assay with lactate dehydrogenase and NADH, as previously described. The assay was performed at 25 °C in a 1-mL cuvette containing 50 mM Tris·HCl (pH 7.5), 0.2 mM NADH, 0.5 U of lactate dehydrogenase, and a suitable aliquot of NAL (1–10 μg). The reaction was initiated by the addition of varying concentrations of substrate (0.5–20 mM Neu5Ac and 0.1–2.5 mM DPAH, chosen to bracket the measured Kₘ values). The decrease in absorbance at 340 nm was recorded on a UVikon 930 spectrophotometer as the measure of enzyme activity. One unit of aldolase activity is defined as the amount of enzyme that catalyzes the oxidation of 1 μmol min⁻¹ NADH in this system, using the molar extinction coefficient of NADH (6220 M⁻¹ cm⁻1). The triol, THB, was synthesized as previously described, and inhibition of the NAL reaction was measured by including THB (0.25–2 mM) in the standard enzyme assay. Steady-state kinetic parameters kₐₑₜ, Kₘ, and Kₛ were all estimated by nonlinear regression to the appropriate rate equations.

Construction of the E192X library

Preparation of the E192X saturation mutagenesis library was undertaken using megaprimer PCR, as previously described. DNA sequencing of 96 mutants revealed 15 of 20 different amino acids at position 192. The remaining five mutants were constructed using QuickChange site-directed mutagenesis, following the instructions provided by the manufacturer (Stratagene).

Protein crystallization

Tests with the pre-crystallization assay kit (Hampton Research) indicated an optimum NAL concentration of 10 mg mL⁻¹. The initial conditions for crystallization were identified with an Oryx crystallization robot (Douglas Instruments) using commercial sparse matrix screens (Hampton Research). Crystals were grown at 4 °C by sitting drop diffusion at a ratio of 2 μL of protein (10 μg mL⁻¹) to 2 μL of mother liquor. Several hits were obtained, but high-salt conditions (such as 3 M (NH₄)₂SO₄ or 2 M Li₂SO₄) were ignored on the basis of our previous studies and the data reported in the literature. The final optimized crystallization conditions were 100 mM Tris–HCl (pH 8.0–8.30), 200 mM NaCl, and 18% (wt/vol) polyethylene glycol (PEG) 3350, which yielded crystals in 7–10 days. This condition is more physiologically relevant.
than the published high-salt conditions\(^6\) and, crucially, it allowed us to soak the substrate pyruvate and the triol inhibitor in the crystals.

The crystals belonged to space group \(P2_1\) with unit cell dimensions \(a = 57\ \AA, b = 143\ \AA, c = 83\ \AA,\) and \(\beta = 109^\circ\) (decimal numbers are omitted due to their variability in several structures). Occasionally, crystals also grew in space group \(P2_1\_2_1_2_1\), and we have recently reported the structure of the E192N variant in complex with pyruvate\(^{19}\) in this crystal form. To form the NAL and the E192N variant pyruvate complexes, we soaked crystals in mother liquor supplemented with 100 mM sodium pyruvate (Sigma Aldrich) and 15% (vol/vol) PEG 400 (Hampton Research) for 1 min. The crystals were then sequentially transferred into mother liquid containing 20% (vol/vol) PEG 400 (with 5% increments in PEG 400 concentration) and finally soaked in mother liquor containing 25% (vol/vol) PEG 400 before being flash cooled in liquid nitrogen. In the case of the pyruvate and inhibitor complexes, the same sequential procedure was performed with the addition of 70 mM THB inhibitor in the last cryoprotection step [25% (vol/vol) PEG 400] for 2 min before flash cooling.

Data collection and structure solution

Diffraction data for each structure were collected from single crystals on the macromolecular crystallography beamlines of Diamond Light Source (stations I02, I03, and I04) at 100 K. The data were indexed, integrated, and scaled using MOSFLM\(^{27}\) and SCALA,\(^{28}\) and further processed to generate a complete list of structure factors with \(R\)\(_{\text{free}}\) flags using the CCP4 suite (Table 2).\(^{29}\) Five percent of the data were randomly selected and excluded from the refinement to constitute the \(R\)\(_{\text{free}}\) set. We have recently reported the crystal structure of the E192N variant in complex with pyruvate in the space group \(P2_1\_2_1_2_1\) (PDB code 2WKJ)\(^{19}\) at a relatively high resolution (1.45 Å). This structure was used as a molecular replacement model to solve the structure of the NAL complex with pyruvate (PDB code 2WNN) in space group \(P2_1\) using Phaser.\(^{30}\) For the other structures presented here (all belonging to space group \(P2_1\)), the initial phases were calculated by combining the initial model obtained from this molecular replacement and the relevant recorded amplitudes after five cycles of rigid-body refinement in REFMACS\(^31\) (treating the entire tetramer as a rigid body).

Refinement

Refinement was performed using REFMACS\(^31\) (Table 2). Isotropic \(B\)-factors were refined in all structures and analyzed using the CCP4 suite program BAVEARGER.\(^{29}\) Riding hydrogen atoms were included in the refinement. After each cycle of refinement, model building was carried out using Coot.\(^{32}\) Coordinates and restraint library files for the ligand THB and for a modified lysine residue, KPI (with a covalently bound pyruvoyl moiety), were obtained from the PRODRG server.\(^{33}\) The quality of the model was checked with MolProbity.\(^{34}\) The superposition of the protein structures and the calculation of their RMSDs were performed with LSQKAB.\(^{29}\)

In all structures, except for that of the E192N variant in complex with pyruvate, REFMAC detected a twinning fraction between 6% and 46%, with a twinning operator \(-h, -k, h+l\). In these structures, refinement was performed throughout using the twin refinement option in REFMAC5.

Accession numbers

Coordinates and structure factors have been deposited in the PDB with accession numbers 2WOS (wild type), 2WNQ (E192N), 2WNB (E192N + pyruvate + THB), 2WNF (E192N + pyruvate), and 2WPB (E192N + pyruvate + THB).

Supplementary materials related to this article can be found online a [doI:10.1016/j.jmb.2010.08.008](http://doi.org/10.1016/j.jmb.2010.08.008)

Acknowledgements

We would like to thank Garib Murshudov for discussions and for help with resolving the high degree of twinning in some of our crystals. The X-ray diffraction data described were collected with the support of Diamond Light Source beamlines I02, I03, and I04. A.R.P. was supported by a Research Councils UK Academic Research Fellowship. We also thank the Biotechnology and Biological Sciences Research Council and The Wellcome Trust for funding.

References


